ChinaXiv合作期刊 图书情报工作
第63卷 第23期 2019年12月

基于专利引文分析的领域科技关联发展态势探究*
——以 HCV 为例

齐燕
中国医学科学院医学信息研究所 北京 100020

摘要：[目的/意义]对科学研究成果与专利技术创新之间知识扩散和交叉融合发展态势的剖析，是学科领域布局、研发投入及科技资源配置等工作的重要决策依据。现有研究的测度指标未能详尽反映科技关联关系的特征及演变规律，通过设计新的指标，提出聚焦演变态势分析的领域科技关联关系测度指标及方法。[**方法/过程**]通过对现有测度指标进行分析归纳，将单一的科学关联度和/或关联时滞指标扩展为紧密性、时效性、多样性三个方面分层次的测量指标，从而较全面地揭示和发现科技关联的程度及特征规律。[**结果/结论**]利用本文构建的测度指标对 HCV 研究领域进行了计算和讨论，揭示出了该领域科技关联发展态势，并体现了新指标在测度维度系统性上的优势以及其结论对专利技术研发活动的参考价值。

关键词：科学关联度 引文分析 科学技术关联发展演变 HCV
分类号：G203
DOI:10.13266/j.issn.0252-3116.2019.23.011

1 引言

随着当前科学技术一体化进程的加快，科学与技术之间的关系已经由相互分离转为相互作用、相互结合、相互渗透、相互转化[1]；科学在对技术产生多方面影响的同时也会受技术的影响与制约[2]。早在20世纪60年代，D. J. Price 教授[3]指出科学与技术有独特的知识积累结构，在一定条件下知识可能从科学流向技术也可能从技术流向科学。之后，诸多学者[4-7]对此进行了更为深入的研究和论证，并立足不同视角，采用调研考察、文献计量学等方法开展了不同学科领域或国家地区的“科学-技术关系”的评估与分析。比如从研究公共科研机构与企业的合作研发活动、追踪人才在科学和技术部门间的流动等角度进行的研究[8-9]；或是借鉴统计学和计量经济模型进行的研究[10-11]；更多的研究则是将学术论文和专利分别视为科学研究成果和技术创新成果的表现形式，采用共现分析及引文分析等科学计量学方法挖掘论文专利之间的关系来开展科技关联关系研究。

一方面，基础研究是应用研究的先决条件和催化剂，是技术创新的重要驱动力[12]；同时，越来越多的专利直接引用科学论文或专著的现象也表明科学和技术的联系日益紧密，知识流动越来越直接与频繁。因此，对科学研究成果与专利技术创新间的知识扩散和交叉融合发展态势的剖析，成为学科领域布局、R&D 研发投入及科技资源配置等工作的重要决策依据。另一方面，在基于专利的引文分析这一主流方法中，现有指标及其运用尚不能细致反映科技关联的一些特征及规律，因此需要尝试进行一些改进工作，比如对现有指标进行细微调整以及设计新的指标等。生物医药领域的专利发明与基础创新之间的关联关系已经有一些较为宏观的论证和剖析，本研究聚焦 HCV（Hepatitis C Virus）研究领域，运用新的指标和方法分析其近十年的科技关联发展态势，以期为该领域的科研活动及科技发展规划管理等工作提供参考，并为其他医学领域的科技关联发展态势研究提供借鉴。

2 相关研究概述

基于专利的论文引文分析法，是通过分析专利对

* 本文系中国医学科学院中央级公益性科研院所基本科研业务费项目“基于论文和专利主题分析的科学-技术关联程度计量方法研究”（项目编号:2017PT63008）和中国医学科学院2017年度协和青年科研基金项目“一种基于创新链理论的协同创新主体遴选方法探究”（项目编号:2017330008）研究成果之一。
作者简介：齐燕（ORCID:0000-0001-9054-5597），助理研究员，博士，E-mail:qi.yan@imicams.ac.cn。
收稿日期:2018-12-23 修回日期:2019-06-13 本文起止页码:97-105 本文责任编辑:杜杏叶

期刊论文、会议论文等非专利型引文(non patent references, NPRs)的引用特征和规律,显性和定量地揭示基础科学研究对技术创新的推动作用。F. Narin 等[13-14]开创了专利计量学测度科学与技术之间的交互关系的研究方法,得到了其他学者[15-16]的推广使用。虽然专利不是技术创新的全部,但是可以作为技术创新的核心内容和基础,对专利引文中的非专利引文的分析可以反映出技术创新对基础研究的依赖程度或者说基础研究对技术创新的贡献大小[17]。M. Meyer[8]的研究也进一步论证了专利对学术出版物的引用是一种重要的衡量科学和技术间交互作用的指标,不同的领域关联强弱、互动方式不同,知识转移机制不同,相应的科技政策也应有所不同。

相比于其他科技关联分析方法,该类方法的研究成果较多。比如,M. Meyer[7]、R. Henderson[18]和 E. Bacchiocchi[19]等分别探测了纳米科技领域的科学－技术关联、美国范围内的科学研究－技术创新关联、欧美主要国家公共科研机构与技术创新企业间的知识关联,以专利对论文的引用关系为"纽带",探测二者在内容、方向、结构、水平上的对应关系。T. Hiroyuki 对日本高影响力专利文献进行科学关联性分析,发现大学的基础研究是发明专利的主要理论来源[20];S CS. Lo 对遗传工程领域的非专利文献引用情况的分析也表明了公共研究机构的基础研究对技术发展的重要作用[21]。科学关联度 SL(Science Linkage)指标是一个常用的指标,国际上有据此考察国家专利技术和基础科学间的关系或影响强弱,比如日本文部科学省的《科技进步年度报告》[22]对比剖析了美国、英国、德国、法国和日本各国的科学关联度这一指标的变化情况。赵志耘[23]等对我国在美国申请的生物科技领域的专利数据进行了研究,论证了我国此领域的科学关联度的增长情况、被引文献的期刊来源、作者国家、机构类型等。裴云龙等[24]使用非专利引文分析方法和负二项回归模型,定量研究了纳米科学对纳米技术的影响;文晓芬[25]同时对化学、医药、计算机与通信、机械四个技术领域开展了技术创新与基础研究的关联关系实证研究。张玲玲等[26]选取 USPTO 的人造纤维技术领域的授权专利,基于专利对论文的引用构建了论文所属期刊的共被引网络,发现了影响该技术发展的重要基础学科,并概括了期刊所表征的基础学科知识对技术创新的影响特征。

纵观上述文献及其他相关研究,经过学者们的理论探究和实证应用,利用专利引文分析展示科学技术关联关系的方法获得了快速的发展,目前学者们所使用的主要计量指标包括篇均非专利引文量、科学关联度、专利引用论文时滞、专利引用论文国家分布、专利引用论文种类分布等。基于此,本研究选取部分指标进行一些调整并借鉴 SL 思想设计新的指标,以期更为细致全面地揭示领域科技关联发展态势和规律,为未来的科技创新活动以及科研管理工作等提供支撑和决策参考,比如提示发明人可能的关键科学知识来源、提示科研活动管理者需要重视的基础理论研究工作等。

3 科技关联关系的测度指标

目前基于专利引文测度科技关联关系的指标主要有科学关联度、关联时滞以及对涉及的期刊的统计分析等。本文将科技关联关系的测度划分为三个层次或维度,定义为紧密性、时效性和多样性。紧密性从数量上表征某特定时段的专利技术与相应基础研究的总体紧密程度及变化趋势;时效性揭示专利技术是否紧跟论文对应科学研究创新进程以及程度;多样性衡量专利技术参考对应科学知识来源的多样性。依据这样的认识,基于现有研究构建本文指标体系设计框架如表 1 所示,下面分别对相应的具体测度指标进行改进设计和论证。指标改进和设计的基本原则是秉承延续科学关联度(SL)指标的简单易用性,运用相对简单的数据统计思想和算式,能够在一定程度上反映考察目标的同时具有推广应用的潜力和价值。需要指出的是,有些维度含有多个指标,可挑选使用。

表 1 科学技术关联测度指标体系框架

维度	指标	意义
紧密性	科学关联度(分时)	从数量上表征某时段的技术与相应基础研究的总体紧密程度及变化趋势
时效性	关联总体(或平均)时滞、基于时滞区间的关联分布	判定专利技术是否紧跟论文对应科学研究创新进程以及程度,可以反映科学知识向技术研发领域的传播速度及各时段知识的应用价值大小,或技术领域吸收科学研究知识的进度和时间跨度
多样性	关联期刊种类、关联学科种类、按文献占比的期刊排序、按文献占比的学科类别排序	衡量专利技术参考相应科学知识来源的多样性;可判定专利技术依赖的核心科学知识来源及对应的重要期刊及学科类别,供专利研发人员参考

3.1 紧密性:科学关联度

CHI 学派将科学关联度定义为专利引用科学文献(包括学术论文、会议论文)的平均数量[27]。为了开展趋势分析,需要进行分时段考察,比如年度。为此,本文将科学关联度公式进行改进,定义为:

$$SL=\frac{\text{该时段专利引用的非专利参考文献的数量(NPR)}}{\text{某时段专利数量(P)}}$$

公式(1)

得到某时段的 SL 数值之后,将多个连续时段的科学关联度数值绘制在一条曲线上即可发现由这些专利所代表的技术与非专利参考文献所代表的科学间关联程度随时间演变的趋势,或许是增大、或许是减少,或许是波浪式。SL 数值越大,表明技术对科学依赖程度越高,或者是有更多的科学知识应用于或流向技术研发和创新,进而创造出新的技术知识。需要指出的是,有学者[28]认为"这个指标被用来考察技术与最新科技发展的关联程度,该数量大,说明研发活动和技术创新紧跟最新科技的发展",本文认为,在考虑时间维度的情况下,这一观点才是适宜的。SL 可从数量上表征(某时段)的技术与相应基础研究的总体紧密程度,反映技术创新对基础研究的依赖程度或基础研究对技术创新的贡献,而是否紧跟最新科技需要从关联时滞指标上评判。

3.2 时效性:关联时滞

理论上,科技关联时滞包含两个方面,一是"科学界"的科学知识流向"技术界"产生技术新知识所需要的时间,二是"技术界"的信息作用于"科学界"产生科学新知识所需要的时间。时滞短,可以说明二者交流互动的速度快,体现出科技创新活动的活跃度很高,当然这也与不同学科或技术领域本身的特性有一定的关系。时滞长,一方面可以反映出二者交流互动速度相对较缓,另一方面也反映出早先的科学知识或技术知识历经长久依然具有适用性和实用性,体现出这些知识的宝贵价值。本文首先关注技术专利引用科学文献的时滞,引用关系是科技知识流动的显性表征,因此这一时滞可以在一定程度上反映科学支撑技术发展的蓄势待发时间。

对技术专利引用科学文献的时滞分析时,可以计算特定专利集合的总体时滞情况,也可以基于科学文献的出版年份分成不同时段考察。时段的划分可以采用相同切分比如每 5 年一段,也可以依据考察目的进行非统一标准的设定,特别是考虑到知识的更新迭代可能年份较新的论文会更多地被引用,所以在需要查看对较新发表的论文的引用情况时可以把较近的年份进行更进一步的拆分和细化,而把较久远的年份进行合并,比如考察专利集合引用的当年的科学知识(t_0)、前两年的科学知识(t_{1-2})、前 3－5 年的科学知识(t_{3-5})、前 6－10 年的科学知识(t_{6-10})、前 11－20 年的科学知识(t_{11-20})、20 年前的科学知识(t_{21-})等,进而统计不同时间区间的科学文献的数量及所占比例,可据此判定相关的技术是否紧跟科学创新进程。这里给出的时段划分仅是一种示例,一方面是为了考察较新发表论文的科学知识被引用的情况或称这些知识的传播情况,同时也是为了表征运用这一指标时可根据实际考察需求来灵活设定不同的区间。计算特定专利集合的总体时滞的公式定义为:

$$P_i\text{年专利引用科学文献时滞(LTD)}=\text{专利年代}P_i-\frac{\sum_{j=1}^{n}\text{科学文献 j 的年代}P_j}{n}$$

公式(2)

其中,n 为 P_i 年专利所引用的科学文献的总数量,P_j 为每篇文献 j 对应的出版时间。这里是基于本文考察某领域不同年份的科技关联时滞的目的设计的公式,因此在某特定年代 P_i 时,专利年代就是 P_i 一个数值;如果是不限于一个年份的专利集合的情况,则需要对多个专利年份取均值或采用其他处理方式获得一个年份数值。

分区间统计被引科学文献的数量与比例的公式表示为:

某时段(TZ_i)的被引科学文献的数量

$$V_i=\sum_{j=1}^{n}X_j \quad \begin{cases} X_j=1, P_j\in TZ_i \\ X_j=0, P_j\notin TZ_i \end{cases}$$

公式(3)

$$\text{某时段}(TZ_i)\text{的被引科学文献的比例}\ PRO_i=\frac{V_i}{n}$$

公式(4)

其中,TZ_i 为第 i 个时间区间,n 为基于科学文献的出版时间划分的所有时间区间的科学文献的总数量,本文给出的各参考时间区间界定如表 2 所示:

表 2 时间区间界定及示例

时间区间标识	专利年度	科学文献出版年限	示例年份(P_i=2015)
t_0	P_i	P_i	2015
t_{1-2}	P_i	P_{i-1},P_{i-2}	2014,2013
t_{3-5}	P_i	P_{i-3},P_{i-4},P_{i-5}	2012,2011,2010
t_{6-10}	P_i	P_{i-6}, P_{i-7}, P_{i-8}, P_{i-9}, P_{i-10}	2009, 2008, 2007, 2006, 2005
t_{11-20}	P_i	P_{i-11}, P_{i-12}, P_{i-13},……	2004,2003,2002,……
t_{21-}	P_i	P_{i-21}, P_{i-22}, P_{i-23},……	1994,1993,1992,……

需要特别指出的是,现有研究中使用的专利年度有的学者是采用专利的优先权申请年;考虑到科学文献从投稿至见刊会有数月的时间间隔,而专利从申请到公开也有延时,所以如果在公式中科学文献的年份

使用的是公开年，那么专利年度使用专利的公开年可能相对更为合理，这样的处理方式是同时考虑了两类创新成果的实际情况并将其放置在了同一标准上。

将参考文献的年份划分为不同区间之后，通过比较在不同区间上的数量或比例，即可看出特定时段的专利技术所参考借鉴的科学知识是较新的多还是较早些年的更多，例如，图 1 所示，t_{1-2}时间区间的科学文献数量最多占比最大，其次是 t_{3-5}时间区间。这里的时效应该更多的是与技术本身的特质有关，放眼技术的整个生命周期，不同阶段参考科学知识的时效特征或许也会有所不同，需要具体情况具体分析。



图 1　不同时间区间参考文献数量及比例示意

3.3　多样性：来源期刊和学科

通常，专利参考的科学文献中期刊论文占据绝大多数，书籍、报告等其他类型相对较少，同时由于现有数据库中期刊文献的信息更为详尽和规范且有较为成熟的分类体系，所以在测度科技关联的多样性时仅采用期刊论文不会影响计算的结论（书籍和报告类文献很少，可以忽略不计）。为此，本文定义基于专利所引用的科学文献所属期刊类别的科技关联多样度（JD）为：

$$JD = \frac{\text{该时段专利引用的论文所属期刊数量}}{\text{某时段专利数量(P)}} * \lambda$$

公式（5）

通常，在 Web of Science（WoS）等数据库中，期刊论文会依据特定的标准被界定为一个或多个学科类别，许多不同的期刊也会属于同一学科，所以可以考虑再拔高一个层级，基于论文所属学科类别（比如 SCI 学科分类）来测度科技关联的多样度，公式如下所示：

$$SD = \frac{\text{该时段专利引用的论文所属学科类别数量}}{\text{某时段专利数量(P)}} * \lambda$$

公式（6）

在两个式子中，λ 作为一个常数用来调整 JD 和 SD 的取值空间，以利于后续分析或进一步的运算，也可为 1。需要特别说明的是，这里分母使用的是某时段的专利总体数量，而不是参考了非专利文献（NPR）的部分专利的数量，是由于旨在与经典的科学关联度保持相同的基准以成为较好的补充指标。这是从技术领域整体角度考察关联情况的思路，如果分母仅考虑有 NPR 的部分专利，则是仅聚焦关联对象本身较为狭义的思路，结果的含义可理解为从篇均关联期刊数量和篇均关联学科数量体现的关联多样度。在专利数量相同或相近的情况下，所参考的文献所属期刊或学科类别越多，JD 和 SD 的数值越大，表明专利技术研发所参考的科学知识来源越广泛。与关联时滞指标类似，对来源期刊和学科的考察也可以从总体种类及各类别比例上开展分析工作。

统计各期刊或者各学科类别的数量可得出占较大比例的来源期刊或学科，可视为专利技术领域的核心科学基础，它们可为该领域的研发人员在后续技术创新过程中查找、吸收和利用。所以，利用 JD 对多样性的测度有助于筛选与专利技术研发紧密相关的科学知识来源的核心期刊、识别某时段的新进期刊；而 SD 的测度则有助于识别关联的核心学科领域和交叉领域。界定这些 TOP 期刊或类别既可以从完整数据集中进行统计运算，也可以分时段考察以发现可能存在的变化或规律。

4　HCV 领域科技关联发展态势分析

4.1　数据来源

丙型病毒性肝炎，简称为丙型肝炎、丙肝，是一种由丙型肝炎病毒（HCV）感染引起的病毒性肝炎，主要经输血、针刺、吸毒等传播，据世界卫生组织统计，全球 HCV 的感染率约为 3%，估计约 1.8 亿人感染了 HCV，每年新发丙型肝炎病例约 3.5 万例[29-30]。丙型肝炎呈全球性流行，可导致肝脏慢性炎症坏死和纤维化，部分患者可发展为肝硬化甚至肝细胞癌（Hepatocell ular carcinoma，HCC）。未来 20 年内与 HCV 感染相关的死亡率（肝衰竭及肝细胞癌导致的死亡）将继续增加，对患者的健康和生命危害极大，已成为严重的社会和公共卫生问题[31]。调研发现，丙肝研究领域专利对论文的引用数量较多。另外，从 WOS 数据库论文数量及被引频次等指标看，该研究领域在近几年都有较高的热度和持续性[32]。因此，本研究选择该领域进行科技关联态势分析，以期通过对该领域科技关联发展态势的分析，为该领域的创新发展提供参考。

专利数据采集选取 DII（Derwent Innovations Index）数据库，检索式为 TS = hepatitis COR HCV OR BCV，年份设定为 2008 –2017，经过遴选，获取了 6662 项专利。其中有3 767项专利参考了科学论文或图书，累计有 30 854条科学文献被引记录；专利的非专利引文相关数据来源自 Lens[33]，检索日期为 2018 年 7 月 20 日。由于下载数据存在信息缺失和错误问题，运用自编程序对下载数据进行了补充和清洗，并进行了人工校对。

4.2 分析结果

4.2.1 关联紧密性 根据式(1)计算从 2008 年至 2017 年公开的 HCV 研究领域专利技术每年的科学关联度,并与每年的专利数量、有 NPR 的专利数量等指标进行对比如图 2 所示:

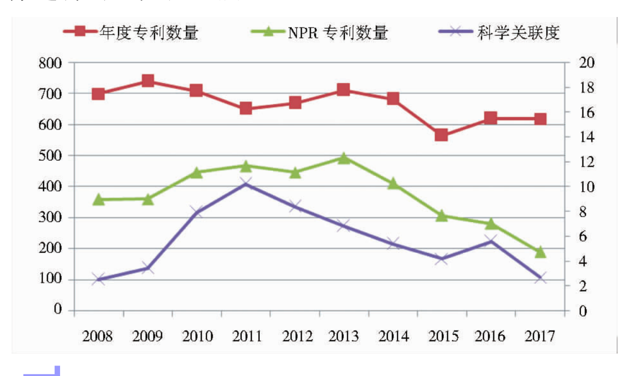


图 2 专利数量及科学关联度趋势

可以看出,近十年专利数量呈现较为平缓的波浪趋势,除 2015 年外,其余年份都在 600 以上;参考了科学文献的专利数量在 2013 年之前呈现递增趋势,之后则逐渐降低,越来越少;相应地,科学关联度也是自 2008 年起递增,至 2011 年达到峰值后逐年减少,仅在 2016 年有一个回升点,总体而言 HCV 研究领域近十年专利技术与科学文献的紧密程度呈现递增转为递减的趋势。

2011 年公开了 651 项专利,其中有 467 项参考了 6 648篇/部科学文献,平均每项引用 14 篇/部科学文献;这一年公开的专利,无论是科学文献类型的参考文献还是专利类型的参考文献,都是十年当中最多的,可以说明这些专利的发明人在此阶段对科学技术知识的学习吸收活动特别活跃。另外值得特别指出的是,2016 年与 2017 年的科学关联度虽然不高,但是这两年公开的专利中所引用的科学文献数量都超过专利类型的参考文献至少 25%,说明发明人们更倾向于从科学知识中获得灵感、启示与借鉴,未来该领域的科学关联度可能会有所增加。

4.2.2 关联时效性 首先根据公式(2)计算各个年度的专利子集对应的总体关联时滞(LTD),结果见表 3。

从关联时滞数值看,2008 年至 2017 年间 HCV 研究领域科学技术关联时滞呈现平缓增长趋势,2016 年最长为 14 年,表明科学知识流向技术研发进而产生技术新知识的速度有所变缓,特别是在最早出版年呈现由远及近、关联年限越来越少的趋势下,可以由此推断近年来在此领域的技术创新难度有所增加。

十年间,最早的平均出版年是 1998 年,最晚的平均出版年是 2004 年,2009－2012 年公开的专利所参考

表 3 年度关联时滞及科学文献出版年份特征明细

专利公开年	参考科学文献出版年均值	参考科学文献出版年中位数	最早出版年	关联年限	关联时滞 LTD
2008	1998	2001	1899	109	10
2009	1999	2001	1898	111	10
2010	1999	2001	1884	126	11
2011	1999	2001	1908	103	12
2012	1999	2002	1906	106	13
2013	2001	2004	1907	106	12
2014	2001	2004	1924	90	13
2015	2002	2005	1929	86	13
2016	2002	2004	1922	94	14
2017	2004	2007	1954	63	13

的科学文献的平均出版年都是 1999 年,其余年份与其相差在 5 年以内,相对较为稳定。出版年的中位数在前几年都是 2001,后几年多是 2004,2017 年的是 2007;最早出版年在前几年大多是在 1900 年前后,后几年则多是 1920 年以后,2017 年的是 1954;关联年限早些年间都是在 100 年以上,后几年逐渐缩短,2017 年为 63 年,即 2017 年公开的专利中所参考的科学文献的出版年份从 1954 年至 2017 年,跨越了 63 个年份。从最早出版年和关联年限的变化上可以看出,随着时间的推移,在对技术研发的支撑上,旧的科学知识逐渐被新的科学知识更替,年份越早的知识被更替的速度越快;而从平均出版年和中位数的变化看,参考的科学知识更替的速度不及专利技术知识的更新速度,体现了作为基础研究的科学知识所具有的持续性的价值。

其次,依照公式(3)和公式(4)计算分时段科学文献的分布情况,绘制比例图如图 3 所示:

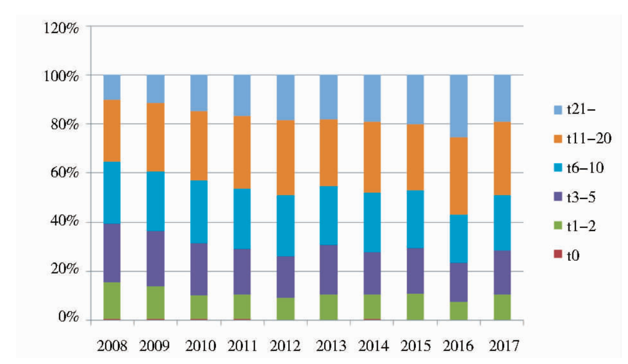


图 3 各年度专利参考的科学文献出版年份分时段比例

总体来看,除 2016 年外,其余年份的专利引用前 0－10 年间出版的科学文献的比例都超过 50%,引用前 20 年以内出版的科学文献比例达到 80%,表明新文献的影响和价值较大;十年来,专利公开前 20 年以前的文献占比逐渐增加,从 2012 年起接近 20%,2016 年时最大,超过 25%,从具体引用数据看,20 世纪 80－90

年代的文献相对更早年间的文献具有更多的份额和更为长久的影响力。

具体地，引用当年出版（t_0）的科学文献占比较少，只有 2008 年的在图中具有一定的可见度，其余都在 0.7% 以下；引用前两年科学文献的比例较为稳定，除 2008、2009 和 2016 年以外，其余都在 10% 左右；2008 至 2010 年以及 2013 年，这四个年度公开的专利参考前 3 - 5 年间出版的科学文献的数量占比都超过 20%，比重较大，其余年份在 15% - 18% 之间；在专利公开年前 6 - 10 年间出版的参考文献最高占到 25%，最低为 19.8%，分别是 2008 年和 2016 年，其余有 4 个年份都是 24% 多，差异不大。

将 t_{11-20} 分成 t_{11-15}、t_{16-20} 两个时段进行细致考察，前五年的占比呈现减少、增加又减少的波浪趋势，相应地后 5 年的占比呈现增加、减少又增加的反向趋势；从体量看，前五年（t_{11-15}）的文献依然是被引用的重点，而从 2014 年起后五年（t_{16-20}）文献的比例有较为明显的增加，说明了这些年份的文献对当今专利技术研发的重要性。

将 t_{1-2} 与 t_{3-5} 两个时段的数据进行合并，与其他时段构成 5 年一个时期进而比较各时期比例可知，各年度公开专利所参考的科学文献中，在前 5 年内出版的占比最多，最大值近 40%，最小值约 24%。其次是前 6 - 10 年，十年间的占比分别比前 5 年内（t_{1-5}）的占比少了 1 - 14 个百分点，其中 2012 年两者仅相差 63 篇，两个五年期均在 25% 上下；差距最大的是 2008 年，t_{6-10} 时段的文献占比相对 t_{1-5} 时段少了 13.7 个百分点，较近的 5 个年度（专利公开年 2013 - 2017）则是在 3 - 6 个百分点左右，差距呈缩小趋势，说明两个时段的文献具有越来越相近的参考价值。前 11 - 15 年的占比在这十年间相差不大，在 15.6% - 18.3% 之间，2011 - 2014 年间数值较大；前 16 - 20 年的占比在 8.8% - 14.4% 之间，较大波动值产生于 2016 和 2017 年，这两个 5 年期的影响力和贡献度相对略小于前两个 5 年期。

总体看来，专利公开前 5 年内的科学文献一直是专利发明人参考的重点，但比重呈平缓缩减趋势，相反，年份较早的科学文献则呈现出平缓增加的趋势，特别是在 2016 和 2017 年，前 16 - 20 年内、前 21 年及更早年份出版的科学文献比重处于十年间的最高或较高水平，未来几年或许会继续保持这一趋势。相对地，前 6 - 10 年内和前 11 - 15 年内的科学文献也是参考的重点，比重较大且随时间推移的变化趋势相对较为平缓，成为较为稳定的科学知识支撑出版年限。这些规律性信息可以作为该领域后续专利发明人进行科学论文或专著类参考文献遴选以吸收科学知识助力专利技术研发时的参考。

4.2.3 关联多样性 对各年度专利参考的每篇科学文献所属期刊、学科类别的信息进行浏览、核查与清洗，进而汇总，按照公式（5）和公式（6）计算关联期刊多样度（JD）和关联学科多样度（SD），并将二者与每年专利数量进行对比如图 4 所示：

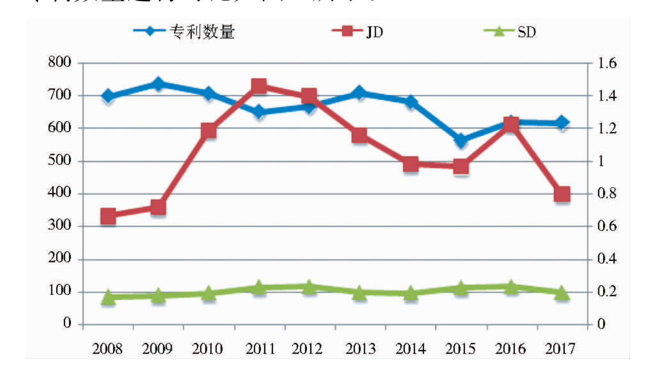


图 4 关联期刊多样度、关联学科多样度趋势

可以看出，学科多样度 SD 的变化趋势较为平缓，值域在 0.2 上下，比专利数量的变化趋势平缓很多，说明支撑这一领域发明专利的基础科学在数量上相对较为固定。其中，10 个年份中有 4 个年份是高于 0.2，最大值是 2016 年约 0.24，最小值是 2008 年的 0.17。期刊多样度 JD 的数值波动较大，值域在 0.67 - 1.46 之间，其中，2010 - 2013 年以及 2016 年的数值较大，主要是由于这几年的发明人参考了较多的科学文献（篇均参考 NPR 数量超过 10 篇，高于其他年份），所涉猎的期刊数量也远多于其他年份，分别是 843、952、936、825、761 种。通常，每种期刊有相对聚焦的主题，期刊种类越多，说明发明人关注和参照的主题数量越多，进而可以一定程度上反映出这几年的发明专利的技术复杂度或者创新程度相对较高。

统计十年来每年的被引量排名前十的（TOP10）期刊，并观察十年分布特征，按照十年累计被引频次列出结果见表 4。

20 种期刊中既有综合性期刊也有专业性期刊，可见科学知识来源的丰富性。表中前两位的期刊十年累计被引频次超过千次，年均超 100 次，堪称重量级参考期刊；第 3 至 9 序号的期刊累计被引频次均超过 300 次；其余的相对较少，但依然可以称为该领域专利技术发明的重要来源期刊，因为每年被引期刊达数百种，绝大部分在十年内只被引用了两次或者一次。所以，表 4 中的期刊可以作为此领域专利发明人后续技术研发

表 4　年度 TOP10 期刊分布特征

序号	期刊名称	进入 TOP10 年份数量	累计被引频次
1	*Proceedings of the National Academy of Sciences of the United States of America*	10	1 284
2	*Journal of Virology*	10	1 095
3	*Journal of Biological Chemistry*	10	751
4	*Nature*	10	713
5	*Journal of Immunology*	10	669
6	*Science*	10	642
7	*Journal of Medicinal Chemistry*	9	451
8	*Hepatology*	7	406
9	*Nucleic Acids Research*	6	310
10	*Vaccine*	3	145
11	*Cell*	2	117
12	*PLOS ONE*	2	69
13	*Bioorganic & Medicinal Chemistry Letters*	2	66
14	*Blood*	2	65
15	*Journal of Experimental Medicine*	1	64
16	*Journal of Hepatology*	1	53
17	*Cancer Research*	2	49
18	*The New England Journal of Medicine*	1	44
19	*Infection and Immunity*	1	36
20	*Virology*	1	20

过程中的首选科学知识来源。值得指出的是,通常情况下,顶级期刊影响大因而单篇文献被引频次往往相对较高,这里统计的是被引频次不是被引文献数量,所以其中一些综合性期刊可能被引用的文献数量不是很多,但是却很值得关注。

具体地,有 6 种(表 4 中序号 1 - 6)在十年中都是排在前十位的重要期刊,特别是 *Proceedings of the National Academy of Sciences of the United States of America*(《美国国家科学院院刊》)和 *Journal of Virology*(《病毒学杂志》)都是位居首位或次位,且被引频次超出后续期刊较多,比如 2009 年二者被引频次分别是 109 和 94,而位居第三的 Journal of Immunology 和 Nature 只有 53 次;其余四种期刊在各年中的位序也比较靠前,特别是 Nature 和 Science 两大国际权威学术刊物的被引频次凸显了这一领域的专利发明人对高质量科学论文的关注,也在一定程度上体现了其科研工作的先进性和高水平。*Journal of Medicinal Chemistry*、*Hepatology*、*Nucleic Acids Research* 三种期刊可以划为第二梯队,分别出现了 9 个、7 个和 6 个年份,其中 *Journal of Medicinal Chemistry* 在 2015 年排名第 12;胃肠肝病学专业期刊 *Hepatology* 在 2009 年排位第 18,在 2016 年和 2017 年分别是 12 和 13;*Nucleic Acids Research* 被引最多的是在 2010 年,位序最高是在 2017 年。表中第 10 - 20 种期刊涵盖了疫苗、病毒、血液、癌症等专题类期刊;相对而言,*Vaccine* 和 *Virology* 在早年间更受关注,而 *Blood* 和 *Cancer Research* 则是在近几年被引较多,体现了作为专利发明人的科研人员研究焦点的转变。

统计十年来每年被引量排名前十的(TOP10)学科类别,并观察十年分布特征,按照十年累计被引频次列出结果如表 5 所示:

表 5　年度 TOP10 学科类别分布特征

序号	学科类别名称	进入 TOP10 年份数量	累计被引频次
1	Molecular Biology	10	4 093
2	Biochemistry	10	4 070
3	Immunology	10	3 655
4	General Medicine	10	3 251
5	Cell Biology	10	2 643
6	Multidisciplinary	9	2 575
7	Molecular Medicine	10	2 547
8	Virology	10	2 504
9	Pharmacology	7	1 794
10	Drug Discovery	6	1 232
11	Microbiology	2	473
12	Genetics	2	455
13	Cancer Research	1	163
14	Infectious Diseases	1	114
15	Immunology and Allergy	1	83
16	Oncology	1	82

16 个学科类别同样具有多样性,从细胞、分子、遗传到病毒、免疫、药理等,还有多学科(Multidisciplinary)这一类别,可以说明这一领域的专利发明人关注到了 HCV 这一领域的很多方面并试图运用多学科知识解决一些更为复杂的问题。分子生物学和生物化学两个类别的被引频次超过 4 000 次,年均 400 多次;表中序号 3 - 10 的类别被引频次都超过 1 000 次,也是该领域的重要科学知识来源学科;其余几个类别相对较少。值得指出的是,有些年份中,排位在 TOP10 以后的学科类别的被引频次与 TOP10 后几位的类别差距不大,比如 2010 年,TOP10 类别为 Microbiology,被引频次有 294 次,TOP11 类别是 Genetics,被引频次是 288 次。

具体地,Molecular Biology、Biochemistry、Immunology、General Medicine、Cell Biology、Molecular Medicine、Virology 等 7 个类别在每年都是入围被引前十,Multidisciplinary 也仅在 2017 年以 1 次之差落后于排位第十的 Virology;Pharmacology 和 Drug Discovery 也出现在了较多年份,前者是 2009 至 2015 年连续入围,后者在

2012 至 2016 表现较突出。后面几个类别入围年份相对较少,但有一定的特征性,比如 2011 年之前是 Infectious Diseases 和 Microbiology,2011 年之后是 Cancer Research、Immunology and Allergy、Oncology,说明近几年学者们对于 HCV 有关的癌症、免疫过敏症等有了较多的关注。

4.3　多指标定性分析

本文给出的三个维度的指标测度了科技关联关系的不同方面,是对科技关联问题进行细化的过程;如果需要归一化的结论则需要对三个维度进行整合,而由于它们具有不同的属性或许不易或不宜进行定量角度的整合,但可以从定性角度对三个维度的结论进行归纳与概括。针对实证领域由上述三个维度的分析可得出,近十年,HCV 研究论文流向专利的知识数量的变化趋势表明该领域科技紧密程度由越来越紧密变得相对疏松,从关联时滞上看科技间知识传递更迭速度逐渐放缓,早年的科学知识依然具有较强的理论或方法支撑价值,参考期刊种类与文献数量变化趋势接近,由越来越广变得逐步收紧,可视为经历了由发散到越来越专注的转变,而参考的学科类别数量则相对较稳定,一定程度上反映出领域发展的稳定性。总体而言,该领域专利技术研发对科学基础研究的依赖相对前几年有一定程度的降低,特别是对新近几年的科学知识,但幅度不大。

5　结语

本文借鉴现有研究的度量指标和方法,以 HCV 领域为例,从演变态势分析的视角探究了揭示医学领域科技关联发展变化的方法。通过对现有成熟指标,比如科学关联度、关联时滞等的改进,包括更为客观的描述、面向趋势分析的细化等,实现了从关联紧密性、关联时效性、关联多样性三个维度多个层面(本文实证了 5 个示例指标:SL、总体时滞、分段时滞、关联期刊、关联学科)来探究领域科技发展过程中知识互动的特征和规律,可为此领域专利发明人在后续技术研发过程中遴选科学知识来源以及科研管理工作者布局研发投入等提供参考。

所设计的新指标虽然相对简单,但是能够反映一些问题,比如从分段时滞的分布上能够看出专利发明人更关注的是哪些年份的科学知识,或者哪些年份的科学知识对专利发明的研发过程起到了更为重要的支撑作用;而从来源期刊或学科的数量、类目和规律上则可以发现重要支撑期刊和研究方向,以及领域发展的稳定性和演变特征,具有一定的实用性和易用性。在实际应用中,可进行更为深入细致的具体文献分析,并可结合其他指标或方法,为科研工作者或科技活动管理者提供更为有价值的信息。需要特别指出的是,Lens 网站为分析专利与其科学文献类型的参考文献提供了诸多便利,但是存在数据缺失及错误信息,需要进行核验和清洗,会花费较多时间。

本文以 HCV 领域进行实证,但指标设计理念没有领域特殊性,所以理论上该方法同样适用于其他学科。未来研究工作将增加实证领域,以技术主题为研究单元,可以对比不同技术主题在各指标上的差异,并能开展多个学科之间的对比。此外,论文的专利引文分析(比如技术关联度)也是测度科技关联的重要方法和思路,所以将二者结合进行双向考察也可能是具有一定意义的探索方向。

参考文献:

[1] 刘则渊,陈悦. 新巴斯德象限:高科技政策的新范式[J]. 管理学报,2007,4(3):346-353.
[2] 刘小玲,谭宗颖,张超星. 国内外"科学-技术关系"研究方法述评——聚焦文献计量方法[J]. 图书情报工作,2015,59(13):142-148.
[3] PRICE D J DE SOLLA. Is technology historically independent of science? a study in statistical historiography[J]. Technology and culture, 1965,6(4):553-568.
[4] VERBEEK A, DEBACKERE K, LUWEL M. Science cited in patents: ageographic"flow"analysis of bibliographic citation patterns in patents[J]. Scientometrics, 2003, 58(2): 241-263.
[5] GARFIELD E. Patent citation indexing and the notions of novelty, similarity and relevance[J]. Essays of an information scientist, 1984, 7(3): 536-542.
[6] NARIN F, HAMILTON K S, OLIVASTRO D. The increasing linkage between US technology and science[J]. Research policy, 1997, 26(3): 317-330.
[7] MEYER M. Does science push technology? patents citing scientific literature[J]. Research policy,2000,29(3):409-434.
[8] BHATTACHARYA S, MEYER M. Large firms and the science-technology interface patents, patent citations, and scientific output of multinational corporations in thin films[J]. Scientometrics, 2003,58(2): 265-279.
[9] CHOUNG J Y, MIN H G, PARK M C. Patterns of knowledge production: the case of information and telecommunication sector in Korea[J]. Scientometrics,2003,58(1): 115-128.
[10] ZHAO Q J, GUAN J C. Modeling the dynamic relation between science and technology in nanotechnology[J]. Scientometrics,2012, 90(2): 561-579.
[11] VAN LOOY B. Do science & technology interactions pay off when developing technology? an exploratory investigation of 10 science-

intensive technology domains[J]. Scientometrics, 2003, 58(3): 355 – 367.

[12] BUSH V. Science and the endless frontier[M]. USA: National Science Foundation, 1945.

[13] CARPENTER MP, NARIN F. Validation study: patent citations as indicators of science and foreign dependence[J]. World patent information, 1983, 5(3):180 – 185.

[14] NARIN F, NOMA E. Is technology becoming science? [J]. Scientometrics, 1985, 7(3/6):369 – 381.

[15] SCHMOCH U. Indicators and the relations between science and technology[J]. Scientometrics, 1997, 38(1):103 – 116.

[16] VERBEEK A, DEBACKERE K, LUWEL M, et, al. Linking science to technology: using bibliographic references in patents to build linkages chemes[J]. Scientometrics, 2002, 54(3):399 – 420.

[17] 赵志耘,雷孝平. 我国生物科技领域技术创新与基础研究关联分析:从专利引文分析的角度[J]. 情报学报, 2012, 31(12): 1283 – 1289.

[18] HENDERSON R, JAFFE A, TRAJTENBERG M. Patent citations and geography of knowledge spillovers: a reassessment: comment [J]. American economic review, 2005, 95 (1):461 – 464.

[19] BACCHIOCCHI E, MONTOBBIO F. Knowledge diffusion from university and public research: a comparison between US, Japan and Europe using patent citations[J]. Journal of technology transfer, 2009 (34):169 – 181.

[20] HIROYUKI T, TAKAYUKI T, YASUHIRO Y, et al. A bibiiometric anaiysis of scientific literatures cited by influential patents [J]. Journal of information processing and management, 2006 (1): 2 – 10.

[21] LO S C S. Scientific linkage of science research and technology development: a case of genetic engineering research [J]. Scientometrics, 2010 (1):109 – 120.

[22] 日本文部科学省. Annual report on the promotion of science and technology 2001[R]. 东京:文部科学省, 2002.

[23] 赵志耘,雷孝平. 我国生物科技领域技术创新与基础研究关联分析:从专利引文分析的角度[J]. 情报学报, 2012, 31(12): 1283 – 1289.

[24] 裴云龙,蔡虹,赵皎卉. 纳米科学对纳米技术的影响基于 NPR 的分析[J]. 情报杂志 2010, 29(10): 1 – 4.

[25] 文晓芬. 基于专利引文的我国基础研究与技术创新的关联分析 [D]. 北京:中国科学技术信息研究所, 2011.

[26] 张玲玲,李睿,李婧. 在期刊 – 专利引用网中发现支撑技术创新的基础知识——以 D01F 技术领域为例[J]. 情报理论与实践, 2014, 37(5) : 64 – 68.

[27] 李春燕,石荣. 专利质量指标评价探索[J]. 现代情报, 2008, 28 (2) :146 – 149.

[28] 殷媛媛. 基于论文专利引证关系的科学技术互动研究——以立体显示为实证分析[J]. 图书情报工作, 2012, 56(16):65 – 70.

[29] DANNY J E, DELPHI G M C, TARA L P, et al. 丙型肝炎药物的可及性[EB/OL]. [2018 – 07 – 10]. https://www.who.int/bulletin/volumes/93/11/15-157784-ab/zh/.

[30] 刘炜炜. 中国 2007 – 2014 年丙型病毒性肝炎新发感染情况估算 [D]. 北京:中国疾控中心流病办, 2016.

[31] 崔妹娟,王晓春. 丙型肝炎流行状况及危险因素的研究进展 [J]. 中国艾滋病性病, 2014, 20(2):141 – 144.

[32] 中国科学院科技战略咨询研究院,中国科学院文献情报中心, Clarivate Analytics. 2016 研究前沿[R/OL]. [2018 – 07 – 10]. http://ipscience.thomsonreuters.com.cn/media/2016researchfront.pdf.

[33] Cambia, Queensland University of Technology. lens[D/OL]. [2018 – 07 – 20]. https://www.lens.org.

Research on the Analytical Method of Science-technology Relationship Evolution in Specific Fields Based on Patent Citation Analysis——Taking HCV as an Example

Qi Yan

Institute of Medical Information/Medical Library, CAMS&PUMC, Beijing 100020

Abstract: [**Purpose/significance**] The analysis of the knowledge diffusion and cross-integration development between scientific research results and patent technology innovation is an important decision-making basis for subject and field layout, R&D investment and technology resource allocation. The metrics of the existing research fail to reflect in detail the characteristics and evolution of the relationship between science and technology, this paper try to design some new indicators, and propose a method for measuring and analyzing the scientific and technological relationship in a specific field that focuses on the analysis of the evolutionary situation. [**Method/process**] By analyzing and summarizing the existing metrics, we expanded the metric of a single indicator such as Science Linkage and/or time lag to the hierarchic almetric that from three aspects: compactness, timeliness and diversity. So that, we can fully reveal and discover the degree and characteristics of science-technology relationship. [**Result/conclusion**] The empirical analysis of HCV research field revealed its development trend of science-technology relationship and verified the systematicness advantage of the new index and the reference value of its conclusion to the patent technology research and development activities.

Keywords: Science Linkage citation analysis science-technology relationship evolution HCV